Data Warehouse Project Overview

2002 IHS Tech Fair
July 9, 2002

Stanley P. Griffith, MD
Stephanie Klepacki
ITSC, Albuquerque, NM

1

Data Warehouse Project

- Why do we need national-level data?
- Why a DW?
- Design features of a DW environment
- Phase 1 – the Pilot Data Warehouse
- Next phase – DW-1

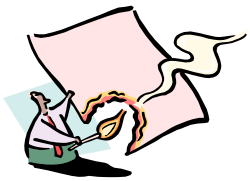
2

Why Do We Need National Data?

- Requests for information are ever increasing
- These ever increasing requests will likely continue
- But our resources will not increase commensurate to these increased demands

3

Why Do We Need National Data?



- By deriving information from existing data, data that are already being collected for other purposes,

we can reduce local data collection burdens and bureaucratic overhead at all levels

4

Why Do We Need National Data?

Information means money

- GPRA is tied to budget requests
- Funding for special needs
  - Racial disparities
  - Diabetes
- ORYX and JCAHO accreditation

5

Why Do We Need National Data?

National level data can improve clinical care

- Patient safety

6

Why Do We Need National Data?

Information provides program direction
- Needs assessment
- Resource allocation
- Outcome performance
- Facility planning

7

Data Warehouse Project

- Why do we need national-level data?
- Why a DW?
- Design features of a DW environment
- Phase 1 – the Pilot Data Warehouse
- Next phase – DW-1

8

How Do We Design Our National DB?

- We have conflicting needs.

9

We Need More Data

The breadth of information needs is increasing
- Statistical
- ORYX
- Epidemiology
- Diabetes
- Pharmacy
- Cost management
- Bioterrorism

etc.

10

We Need One, All-Inclusive Repository

One central collection point for all data
- Verify receipt of data
- Analyze and provide feedback on
  - Timeliness of data
  - Data quality
  - Unexpected deviations from historical norms
- Maintain a "single version of historical truth"
- Maintain all the information content of the data

11

One, All-Inclusive Repository

- And one large information repository becomes too unwieldy for reporting

12

We Need Focused, Efficient DBs

We need efficient, user friendly access to data

- Ease of user access
- Search-efficient DB structures
- Subsets of data – just what we need for specific uses
- Transformed data

13

Focused, Efficient Data Marts

But data marts alone just cannot maintain the

- Flexibility
- Granularity
- Scalability
- Power

that will be required to meet all needs now or in the future

14

We Must Accommodate Growth

And the architecture has to be able to accommodate change.

There will be future needs for information that even the most astute among us cannot now anticipate.

15

We Need Both…

We need one, large, replete, and powerful data warehouse,

that provides information to more focused, user-friendly, efficient data marts.

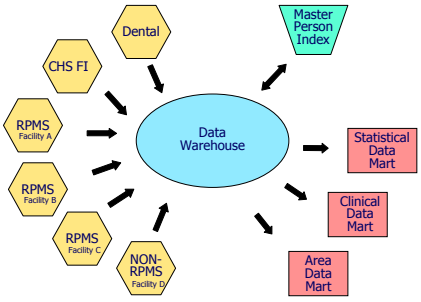
16

Data Warehouse Project

- Why do we need national-level data?
- Why a DW?
- Design features of a DW environment
- Phase 1 – the Pilot Data Warehouse
- Next phase – DW-1

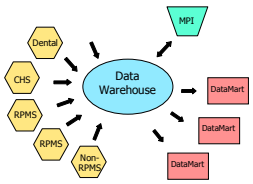
17

National DW Environment



18

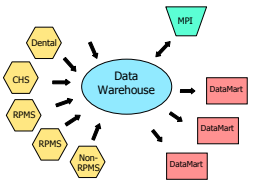
The DW Environment Will



- Be designed to take advantage of the latest in technologies and design features

19

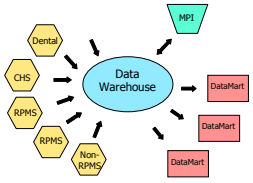
The DW Environment Will



- Collect and store information exactly as it is received from the field

20

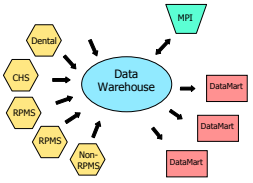
The DW Environment Will



- Provide accurate and intelligent information about the data it receives:
  - Record counts
  - Timeliness of data
  - Less than expected counts based on historical benchmarks
  - Missing data in fields, erroneous codes

21

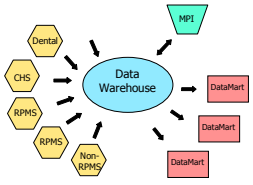
The DW Environment Will



- Supply data to various data marts from which more targeted information can be gathered
  - Statistical - workload and user pop reports
  - Clinical – GPRA, ORYX, Epidemiology
  - Area
  - Pharmacy

22

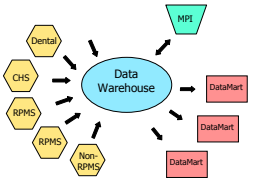
The DW Environment Will



- Employ a Master Person Index (MPI), when available

23

The DW Environment Will



- Employ healthcare industry standards for data transport and messaging

24

Data Warehouse Project

- Why do we need national-level data?
- Why a DW?
- Design features of a DW environment
- Phase 1 – the Pilot Data Warehouse
- Next phase – DW-1

25

PDW: Phase 1

- Three Areas - Albuquerque, Nashville, Phoenix
- RPMS and non-RPMS sites
- Data from FY 97 to present
- Large spectrum of data content
- Working with experts from IBM and SAS

26

PDW: Phase 1

- Designed to gather and store as much granularity as possible
- Minimal data cleansing and data transformation between staging and DW tables
- No unduplication of visit records *yet* if they come from different sources
- Store all modified, deleted records, flagging the current snapshot

27

PDW: Phase 1

- Exploring data in PDW will determine which data transformations should be hard-coded in DW-1 and where
- Vying priorities
  - Creating one single "version of truth" for all DMs
    - Consistency
    - Efficiency
  - Data mart specific transformations

28

PDW Status

- Logical and physical models complete
- Load status
  - ✓ Registration data
  - ✓ PCC encounter data
  - ✓ CHS Fiscal Intermediary data
  - ✓ CHS 638 (non-FI) data
  - ✓ Dental data

29

Testing PDW

- Producing Workload and User Pop Reports
- Performing several clinical outcome measures
- Field content reports
- Deviation from historical norm reports
- Exploring
  - Unduplication methods
    - Patients
    - Visits
  - Handling of receipt of updated records
  - Storing non-uniform data
  - Modeling complex multi-hierarchical and non-hierarchical relationships

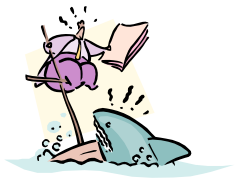
30

Data Warehouse Project

- Why do we need national-level data?
- Why a DW?
- Design features of a DW environment
- Phase 1 – the Pilot Data Warehouse
- Next phase – DW-1

31

The Dilemma…



- We need a complete, all-inclusive solution and we needed it yesterday
- But to plan, build, and implement a complete solution would take far, far too long and would not deliver incremental value to our users along the way

32

The Solution

- Build the data warehouse environment in increments that deliver value to users at each step
- The most critical user needs will be addressed first

33

DW-1

- Extend PDW system-wide
- Export enhancements
- Provide information to data sources about data receipt, integrity, and quality
- Implement a clinical and statistical data mart
- Provide user access to data marts
- Gather and document metadata

34

Extend PDW System-Wide

- Enhance PDW architecture and processes to improve performance and functionality and support future expansion
- Adapt PDW models to accommodate limited number of additional data elements and different entity relationships
- Enhance data cleansing and transformation routines based on what we learned from PDW

35

Extend PDW System-Wide

- Key issue: scalability

36

Export Enhancements

- HL7 based export from RPMS sites
- Unified RPMS registration/encounter export
- Utilize Interface Engine

37

Data Integrity and Quality

- Provide information to data sources
  - Export tracking
  - Error checking
  - Data quality
- Combine elements of
  - ORYX data tracking system
  - PDW web-based reports
  - PCC encounter export monitoring project

38

Data Marts

- Two initial data marts
  - Statistical – "son of NPIRS"
  - Clinical – "daughter of ORYX"
- Near future data marts
  - Data quality
  - Pharmacy
- Not quite as "near future"
  - Area
  - Dental

39

User Access

- Secure, ad hoc access for authorized users to appropriate data marts
- Awaiting approval for draft guidance on access and use of data

40

The Future for Data Marts

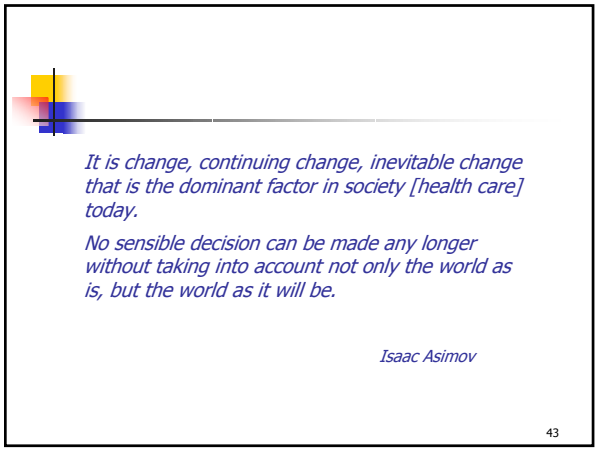
- Data Marts will multiply and ultimately be planned, built, and maintained by programs rather than IT
- Information and the ability to derive useful knowledge from it will become even more critical to our survival

41

Metadata

- Data about data
- Collected and stored in DB2 Warehouse Manager
- Eventually part of a central metadata repository

42


It is change, continuing change, inevitable change that is the dominant factor in society [health care] today.
No sensible decision can be made any longer without taking into account not only the world as is, but the world as it will be.
Isaac Asimov
43